

AD\_\_\_\_\_

Award Number: DAMD17-03-1-0648

TITLE: MTA1-Regulated Gene Expression: New Markers of Breast  
Cancer Metastasis

PRINCIPAL INVESTIGATOR: Peter O'Connell, Ph.D.

CONTRACTING ORGANIZATION: Virginia Commonwealth University  
Richmond, Virginia 23298-0568

REPORT DATE: September 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20050218 037

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 2004		3. REPORT TYPE AND DATES COVERED Final (18 Aug 2004 - 17 Aug 2004)
4. TITLE AND SUBTITLE MTA1-Regulated Gene Expression: New Markers of Breast Cancer Metastasis			5. FUNDING NUMBERS DAMD17-03-1-0648	
6. AUTHOR(S) Peter O'Connell, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Virginia Commonwealth University Richmond, Virginia 23298-0568  E-Mail: poconnell@vcu.edu			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words)  Doctors examine primary tumors and nearby lymph nodes to assess cancer spread, as node-positive breast tumors require obligatory chemotherapy. Most breast cancers are "node-negative" (confined to the breast), but, if left untreated, one fourth of these patients die from occult systemic disease ("micrometastasis"). Women with node-negative breast tumors that overexpress metastasis-associated 1 (MTA1) protein had recurrence risks identical to women with systemic disease (RR 2.7, p = 0.0006). For the 7% (29/394) of patients with tumors expressing the highest levels of MTA1, relapse rates exceeded 60% (1). MTA1 detected recurrence risk before relapses were clinically evident, MTA1 is a prognostic marker and may herald micrometastasis.  We plan to treat breast cancer cells with antisense and control oligomers with an antisense construct that specifically blocks MTA1 protein formation. Gene expression microarrays can measure levels for nearly every gene. Genes that are differentially expressed in AS-MTA1-treated versus control breast cancer cells are either directly or indirectly MTA1-regulated. Since MTA1 is known to enhance tumor metastasis (2,3), future studies of differentially expressed (MTA1-regulated) genes will confirm their role in primary breast tumor metastasis. MTA1, gene expression arrays, and new micrometastasis markers should allow doctors and patients to objectively weigh the need, and likely clinical benefits of breast cancer chemotherapy.				
14. SUBJECT TERMS Breast cancer			15. NUMBER OF PAGES 26	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

## Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	4
Key Research Accomplishments.....	7
Reportable Outcomes.....	9
Conclusions.....	9
References.....	10
Appendices.....	11

## **Introduction**

The presence or absence of systemic disease is the most crucial factor determining survival versus mortality in women with breast cancer. Identifying high risk women and ensuring they receive appropriate systemic treatment reduces risk of death from breast cancer metastasis. However, other than local cancer spread, tumor size, at proliferation, very few clinically useful prognostic markers exist, especially for lymph negative patients.

We have identified a candidate gene, designated metastasis-associated 1 (MTA1), as a possible prognostic and predictive marker both by genomic analysis, IHC of breast tumors, and gene expression analysis (1, 4). Our aims are to identify MTA1-regulated genes to further assess genetic mechanisms promoting breast cancer spread.

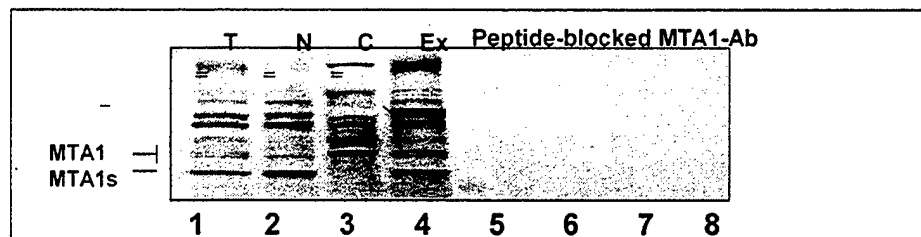
## **Body of Research**

Inferring MTA1's role in metastasis is complicated by a rapidly growing MTA-gene family, so far six published alternatively spliced variants encoded at three genomic loci (i.e., MTA1-14q; MTA2, aka MTA1L1-11q; and MTA3 aka MTA2-2q). Nuclear MTA1 protein functions as a steroid hormone receptor co-repressor (5, 6, ). Multiple alignments of the MTA1, -2, and -3 genes identified an 14q-locus specific peptide which, when linked to a hapten and used to immunize rabbits, yielded a n MTA1-specific polyclonal antibody. Other key reagents for this study were breast cancer cell lines: MCF7, T47D, MDA-MB-435S, and MDA-MB-231 and an antisense MTA1 morpholino-oligomer (AS-MTA1) designed to knock down of MTA1 protein translation. Equipped with these reagents, we 1) studied the cellular distribution of MTA1 isoforms;2) carried out AS-MTA1 transfections and monitored turnover of endogenous MTA1 protein by western blot

analysis. , and 3) analyzed differential gene expression in AS-MTA1 versus control-treated breast tumor cells to identify MTA1-regulated genes.

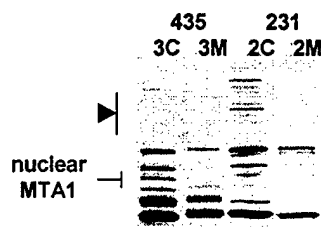
Several observations emerged from these studies.

1. The transfection method (EPEI, GeneTools, Inc) used salts out AS-MTA1 oligomers and carrier DNA in a precipitate that upon cell phagocytosis, oxidizes and chemically bursts endosomes to release the AS-MTA1 oligomer. This technique proved reliable in metastatic MBA-MB-435S and MDA-MB-231 cells, but the MCF7 and T47D cell lines proved completely refractile to AS-MTA1 treatment, as their limited motility may reduce opportunities to accumulate the AS-MTA1/DNA/EPEI precipitate. We focused on the metastatic MDA-MB-231 and -435S cell lines to identify MTA1-regulated genes, and, as an alternative to As-MTA1 in the treatment-resistant MCF7 and T47D cells, plan to may stable transfectants overexpressing MTA1. Figure 1 shows western bolt MTA1 protein is suppressed upon AS-MTA1 treatment.



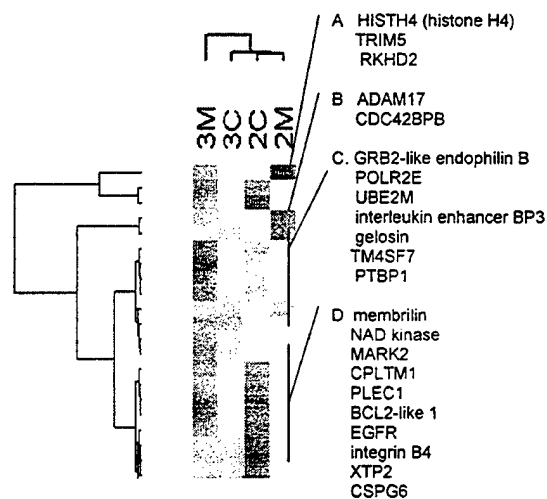
**Figure 1. Western blot analysis of antibody specificity, MTA1 isoforms and their cellular localization.** Lanes 1-4: The MCF7 breast cancer cell line and its MTA1 isoforms (1- total cell lysate, 2- nuclear, 3- cytoplasmic, 4- MCF7 overexpressing pcDNA3.1-MTA1. Lanes 5-8: Western blot with same samples as 1-4, but with the anti-MTA1 antibody preabsorbed to excess MTA1 peptide immunogen to validate specificity.

2. MTA1 undergoes alternative splicing to the nuclear full length MTA1 and a more recently described "short" cytoplasmic isoform, MTA1s. MTA1s shares the full length protein's N-terminus, but replaces the C-terminal SH2-nuclear localization domain with a distinct C-terminal estrogen receptor-binding (LRILL) motif (7).



**Figure 2.** Western Blots of MTA1 in control (2C, 3C) and AS-MTA1 (2M, 3M) treated MDA-MB-435S (3C and 3M), and -231 (2C and 2M). The MTA1-AS treatment suppresses several MTA1 antibody reactive proteins that are differentially expressed between the two breast cancer cell lines.

3. MDA-MB-231 and -435S express differing subsets of MTA1 isoforms, MDA-MB-231 expressed immunoreactive proteins that were AS-MTA1 suppressed. We suspect a number of the other MTA1-antibody immunoreactive proteins (best seen in figure 1, lanes 3-6) result from alternative spliced MTA1 variants that use an upstream initiating methionine that are not targeted by the AS-MTA1 oligomer tested. RNA from treated and untreated cells was hybridized to Affymetrix HG-U133A genechips. These microarrays measure expression of 14,500 annotated genes and 3,900 gene variants.



**Figure 3.** MTA1-knock down associated differences MDA-MB231 cells (2M, 2C). Also shown is expression of the same genes in treated 435S (3M, 3C) cells. Genes in the four MTA1-regulated gene clusters (A-D) are identified (right). AS-MTA1-associated gene expression differences between treated and untreated MDA-MB-231 cells are shown for illustration purposes, as >250 genes showed >2-fold expression difference in treated -435S cells

The DNA Chip Analyzer v1.3 (8, 9) program was used to normalize the array data and model gene expression. Differential gene expression ( $> 2$ -fold changes,  $p < 0.05$ ) AS-MTA1-treated MDA-MB-231 breast cancer cells is summarized in Figure 3, which also shows the behavior of the same genes in the less metastatic MDA-MB-435S (3M, 3C) breast cancer cell line. In MDA-MB-231, AS-MTA1 treatment altered the expression just 22 genes in four clusters (A-D). Gene expression patterns in the treated MDA-MB-231 and -435S cells (2C, 3C) were distinct, while those for the AS-MTA1 treated (2M, 3M) cells converged. The non-overlapping MTA1 isoform repertoires in the two cell lines affect MTA1 regulated gene expression. For example, cluster "A" shows coordinate regulation of the TRIM5, and RKHD ring finger genes. In the -231 line, histone H4, an MTA1 acetylation target (11), is downregulated by AS-MTA1 treatment (i.e., higher expression in the presence of MTA1), but MTA1 upregulated in AS-MTA1 treated MDA-MB-435S. Differences in clusters "B" and "D" clusters were -231 specific, while those in "C" were coordinately regulated in both cell lines. Consistent with likely MTA1 functions, differentially expressed gene's functions include apoptosis (BCL2-like, GRB2-like endothelial B), cell motility/adhesion (ARGIDA, CDC42BPA, CDC42BPB, gelosin, CLPTM1, membralin, ITGB4, MARK2, PLEC1) cell proliferation (ADAM17, EGFR), and gene transcription (ILF3, RKHD2, CDC42BPA, CSG6 RNAPOL2). Our IHC analysis detects a link between MTA1 and AC chemoresponse (1), and microarray analysis of untreated tumor core biopsies showed MTA1, microtubule- and mitotic spindle-associated genes are differentially expressed, suggesting a potential roles in docetaxel chemoresponse.

#### **Key Research Accomplishments**

1. IHC studies of tumor with long term follow up showed MTA1 is a prognostic factor for breast cancer recurrence risk, women whose primary tumors overexpressed nuclear MTA1 protein had a 2.7-fold increase in recurrence risk.

2. The same study indicated MTA1 overexpression associated with adriamycin-cytosine and tamoxifen chemosensitivity. In addition, our report (Chang et al 2003) report found 3-fold higher levels of MTA1 mRNA in docetaxel-sensitive breast tumors. We will continue to study the roles of MTA1 isoforms in breast tumor chemoresponse.
3. Western blot analyses of breast cancer cell lines identified new MTA1 isoforms, and the AS-MTA1 gene expression studies suggest these different isoforms differentially modulate MTA1-regulated gene expression. We plan further studies on how MTA1 isoform heterogeneity affects its performance as a prognostic and predictive factor
4. Work was supported in part by this funding produced a co-transgenic MMTV- MTA1 and MMTV-EGFP (green fluorescent protein) mouse model. These animals are being bred to single oncogene-derived (MYC, Ras, and neu) transgenic mouse models of primary breast tumors. We plan to investigate these animal for de novo MTA1-induced metastases (tg-Ras, Myc tumors do not metastasize), and MTA1-enhanced metastasis (tg-Neu tumors produce spontaneous mets at low frequency).
5. A manuscript:  
"Martin MD, Hilsenbeck S, Mohsin SK, Hopp, TA, Clark GM, Osborne CK , Allred DC, O'Connell P. Breast tumours overexpressing nuclear isoforms of metastasis-associated 1 (MTA1) protein have high recurrence risks and enhanced response to systemic therapies" was submitted to Lancet, but returned without review based on editorial priorities. It will be reformatted and submitted to another journal.
6. A manuscript describing novel isoforms of MTA1, and differential gene expression in AS-MTA1-treated versus control breast cancer cell lines is in preparation
7. Work supported in part by this funding will be reported in a manuscript characterizing a transgenic mouse overexpressing the 80 kilodalton MTA1 steroid hormone co-repressor isoform as well as echinoderm green fluorescent protein in mammary gland (MMTV-promoter



driven) is planned once a sufficient number of animals have been characterized.

### **Reportable Outcomes**

1. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, Mohsin S, Osborne CK, Chamness G, Allred DC, O'Connell P. Gene expression profiling predicts therapeutic response to docetaxel (Taxotere™) in breastcancer patients. Lancet 362:364-369, 2003 (reprint attached as a .pdf file)
2. We have just confirmed a transgenic mouse line co-expressing the MTA1 nuclear steroid hormone co-repressor isoform and green fluorescent protein (to facilitate detection of micrometastases) under control of an MMTV promoter,
3. Data collected under the aegis of this funding led to the award of a "V" Foundation clinical translation grant to the PI (O'Connell).

### **Conclusions**

Concept award (DAMD17-03-1-0648) funding of "MTA1-regulated gene expression: New markers of breast cancer metastasis" permitted my laboratory to make significant progress investigating MTA1 as a prognostic marker of occult systemic disease in node-negative patients, and as a predictive marker of tumor chemoreponse. Our data suggest that MTA1 overexpressing node-negative primary breast tumors are at high-risk for recurrence, and that patients whose locally treated primary tumors show favorable risk marker profiles and normal levels of MTA1 expression might be spared adjuvant chemotherapy. Furthermore, the association between MTA1 overexpression and enhanced chemoresponse has potential implications for all breast cancer patients, and warrants additional study. I greatly appreciate the U.S. Army Medical Research and Materiel Command for their support of my research.

## References:

1. Martin MD, Hilsenbeck S, Mohsin SK, Hopp, TA, Clark GM, Osborne CK, Allred DC, O'Connell P. Breast tumours overexpressing nuclear isoforms of metastasis-associated 1 (MTA1) protein have high recurrence risks and enhanced response to systemic therapies, manuscript in revision (ms. attached to this report).
2. Hofer MD, Menke A, Genze F, Gierschik P, Giehl K. Expression of MTA1 promotes motility and invasiveness of PANC-1 pancreatic carcinoma cells. *Br J Cancer* 90:455-62, 2004
3. Mahoney MG, Simpson A, Jost M, Noe M, Kari C, Pepe D, Choi YW, Uitto J, Rodeck U. Metastasis-associated protein (MTA1) enhances migration, invasion, and anchorage-independent survival of immortalized human keratinocytes. *Oncogene*. 21:2161-70, 2002
4. Martin MD, Fischbach K, Osborne CK, Mohsin SK, Allred DC, O'Connell. LOH Events Impeding Breast Cancer Metastasis Contain the MTA1 Gene. *Cancer Res* 61:3578-80 (2001)
5. Yao YL, Yang WM. The metastasis-associated proteins 1 and 2 form distinct protein complexes with histone deacetylase activity. *J Biol Chem*. 278:42560-8. Epub 2003
6. Talukder AH, Gururaj A, Mishra SK, Vadlamudi RK, Kumar R. Metastasis-associated protein 1 interacts with NRIF3, an estrogen-inducible nuclear receptor coregulator. *Mol Cell Biol*. 24:6581-91, 2004
7. Kumar R, Wang RA, Mazumdar A, Talukder AH, Mandal M, Yang Z, Bagheri-Yarmand R, Sahin A, Hortobagyi G, Adam L, Barnes CJ, Vadlamudi RK. A naturally occurring MTA1 variant sequesters oestrogen receptor-alpha in the cytoplasm. *Nature* 418:654-7, 2002

Breast tumours that overexpress nuclear metastasis-associated 1 (MTA1) protein have high recurrence risks but enhanced responses to systemic therapies.

Michelle D. Martin<sup>1,2</sup>, Susan G. Hilsenbeck<sup>2,3</sup>, Syed K. Mohsin<sup>2,4</sup>, Torsten A. Hopp<sup>2,3</sup>, Gary M. Clark<sup>2,3</sup>, C. Kent Osborne<sup>2,3</sup>, D. Craig Allred<sup>2,4</sup>, and Peter O'Connell<sup>1,2,5\*</sup>.

Deleted:

Deleted: ing

Deleted:

Deleted: isoforms of

Deleted: and

Deleted: response to

Deleted: ies

<sup>1</sup>Department of Molecular and Cellular Biology, <sup>2</sup>Breast Center, <sup>3</sup>Department of Medicine, <sup>4</sup>Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, and <sup>5</sup>Department of Human Genetics, Virginia Commonwealth University, Richmond, Virginia 23298-0033.

Deleted:

Formatted: Superscript

Deleted:

\*Correspondence should be addressed to:

Peter O'Connell, PhD, Department of Human Genetics, VCU School of Medicine, 1101 East Marshall Street, P.O. Box 980033, Richmond, Virginia 23298-0033

Deleted:

Deleted: Department of Human Genetics,

Deleted: Virginia Commonwealth University

Tel: (804) 828-9632, X240

Fax: (804) 828-3760

E-mail: [poconnell@vcu.edu](mailto:poconnell@vcu.edu)

Requests for reprints should be addressed to: Peter O'Connell, Ph.D., Department of Human Genetics, Virginia Commonwealth University, P.O. Box 980033, Richmond, Virginia 23298-0033.

Deleted: ¶

Deleted: ¶  
Report (856 words)

Deleted: (92 words)

#### **(Summary- 100 words)**

Deleted: M

Nuclear metastasis-associated one (MTA1) protein is an oestrogen receptor co-repressor regulating transcription via chromatin remodeling. MTA1 mRNA levels are elevated in metastatic relative to non-metastatic tumours. MTA1 loss of heterozygosity is significantly less frequent in node-positive relative to node-negative breast tumours, suggesting epigenetic alterations of MTA1 affect metastatic potential (1)

Deleted:

Deleted: ing

Deleted: is lower

Immunohistochemistry showed that MTA1 overexpressing tumours have recurrence

Deleted: [Is it acceptable to have a reference in the summary?]

risks similar to node-positive tumours. Untreated node-negative tumours that overexpressed MTA1 had the highest relapse risk (HR = 2.68, p = 0.0006).

Deleted: ing

Chemotherapy eliminated all MTA1 associations with clinical outcome, suggesting

MTA1 overexpression predicts early relapse, but is associated with enhanced chemoresponse.

Deleted: ognoses

Deleted: and

Deleted: s

(Report-860 words)

The prognostic significance of lymph node metastases has long been known, to dichotomize risk of local versus systemic breast cancer, as have steroid hormone receptors' predictive value in selection of adjuvant hormonal therapy versus cytotoxic chemotherapies. While extremely useful, the risk factors in current use have limits. Finding a breast tumour to be oestrogen receptor-negative cannot infer the patient's optimal chemotherapy regimen, and the absence of lymph node metastases cannot stratify relapse risks for node-negative patients. Due to improved awareness and screening programs, women with primary breast cancer increasingly present with node-negative disease. Since the biomarkers in current use cannot differentiate risks for node-negative patients, most opt for chemotherapy, although relatively few stand to benefit. Effective prognostic indicators of micrometastasis could stratify recurrence risks and adjuvant therapy benefits, sparing the majority of these women from the toxicity and cost of chemotherapy.

MTA1 is a steroid hormone receptor co-repressor (2), but inferring a specific role for MTA1 in metastasis is complicated by the rapidly growing MTA-gene family's at least six alternatively spliced forms encoded at three separate loci (i.e., MTA1 at 14q; MTA2 at 11q; MTA3 at 2q). Multiple alignment of MTA gene family open reading frames identified an MTA1-specific peptide that when attached to a hapten, generated a rabbit anti-MTA1 polyclonal antibody. MTA1 undergoes alternative splicing to both full length MTA1 and a previously described "short" cytoplasmic isoform (MTA1s) that replaces full length MTA1's C-terminal src homology and nuclear localization domains with a distinct C-terminal ER-binding (LRILL) motif (2). MTA1s interacts with ER $\alpha$  in cytoplasm rather than the nucleus (2).

We studied a large collection of archived primary breast cancers with an average of 8.8 years of clinical follow-up. Only 15% of the primary breast tumours studied showed significant cytoplasmic immunohistochemical (IHC) staining, but to avoid confusion based on MTA1s' likely alternative function, only nuclear IHC signals were scored for these analyses. MTA1 nuclear IHC signals were scored on a range of 0-8 by adding a five point proportional score for percent of IHC positive cells to a three point IHC staining intensity scale (3). To define MTA1 overexpression, we compared MTA1 nuclear IHC scores measured in normal versus tumour tissues. Breast tumour specimens tested had a significantly higher IHC score (3.57 versus 5.07, respectively, for normal and tumour tissues;  $p < 0.0002$ ). As IHC scores

Deleted: :

Deleted: 56

Deleted: The prognostic significance of I

Deleted: status

Deleted:

Deleted: has long been recognized to

Deleted:

Deleted: however,

Deleted: F

Deleted: Given node-negative patient's 30% recurrence risk, and 30% therapy response [Not sure what this means. We usually think of response as tumor shrinkage, which is not applicable to primary breast cancer. Are you referring to the Overview analysis where response means reduction in the odds of relapse or death?]

Deleted: less than 10%

Deleted: accrue significant clinical

Deleted: (aka MTA1L1)

Deleted: -

Deleted: -

Deleted: MTA1, MTA2, and

Deleted: 3

Deleted: , which shares full length MTA1's N-terminus, but

Deleted: the

Formatted: Font: Italic

Deleted: SH2-

Deleted: ;

Deleted: and

Deleted: ry

Deleted: so

exceeding 5 occurred in less than 5% of normal tissues, we defined MTA1 overexpression as an IHC score equal to or greater than 6. Correlation analyses found no association between MTA1 expression, positive lymph nodes, or tumour size. As shown in **Table 1**, multivariate analysis of the full tumour set revealed that MTA1 overexpression was significantly associated with early relapse (HR = 1.91  $p$  = 0.0015). To avoid bias created by adjuvant endocrine and/or cytotoxic therapies, node-negative patients were separated into treated (N=217) and untreated (N= 397) subsets. In the untreated subset, both univariate and multivariate analysis indicated MTA1 overexpression was a strong prognostic indicator of early disease recurrence (HR = 2.68,  $p$  = 0.0006), outperforming both tumour size (HR = 1.41,  $p$  = 0.039), and S-phase fraction (HR = 1.26,  $p$  = 0.072). As indicated in **Figure 1**, the 23% (93/394) of untreated node-negative patients whose tumours overexpressed MTA1 levels had significantly increased risk of early disease ( $p$  = 0.0001 in univariate analysis and  $p$  = 0.0006 in multivariate analysis). For the 7% (29/394) patients whose tumours expressed the highest levels of MTA1 (IHC score 7-8), relapse rates exceeded 60%. **Table 1** indicates that despite MTA1's nearly 2-fold increase in recurrence risk, neither univariate nor multivariate models of overall survival detected any association between MTA1 overexpression and earlier patient death ( $p$  = 0.42). **Table 1** also shows that the treated subset of node-negative patients had no MTA1-associated increase in recurrence risk ( $p$  = 0.61). The acute risk yet unaffected survival seen in locally treated node-negative patients suggests that upon systemic treatment, their MTA1-overexpressing recurrent disease had enhanced treatment responses. These findings indicate that MTA1 overexpression is an independent prognostic indicator of risk of early relapse, especially in untreated lymph node-negative primary breast cancers. MTA1 overexpression fails to directly associate with robust indicators of recurrence such as tumour size and lymph node status, suggesting that MTA1-facilitated distant spread is independent of, and perhaps distinct from, lymph node-associated recurrence risk. As a result, measurement of MTA1 by IHC gleaned independent information and increased the sensitivity of multivariate models of prognosis. MTA1 retained its prognostic significance in node-negative disease, and dichotomized otherwise unremarkable untreated node-negative primary breast tumours in low versus high risk subsets. Surprising as MTA1 overexpression-associated endocrine and cytotoxic tumour cross-sensitivity to treatment might appear, in an essentially unrelated study that considered neoadjuvant docetaxel response as a function of pre-treatment breast

Deleted:

Deleted: an even

Deleted: er

Deleted: s

Deleted: P

Deleted: s

Deleted:

Deleted: Given that all untreated node-negative patients who recurred received adjuvant therapy [I don't understand; how could untreated patients have received adjuvant therapy?], this observation suggests that MTA1 overexpression is

Deleted: associated with enhanced treatment response.¶  
These

Deleted: suggest

Deleted: -positive, disease

Deleted: metastasis

Deleted: Unlike nodal status,

Deleted: (tamoxifen)

Deleted:

Deleted: (doxorubicin with cytoxan)

tumour gene expression profiles, we found MTA1 mRNA levels were elevated 2.9-fold ( $p = 0.0085$ ) in the docetaxel-sensitive primary tumours (4).

In summary, our data suggest that measuring MTA1 protein expression in primary breast tumours identifies a high-risk subset of node-negative patients who need aggressive treatment, and a larger subset with no MTA1-associated recurrence risks. Furthermore, the strong association between MTA1 overexpression and enhanced treatment response has potential implications for all breast cancer patients, and warrants additional study.

#### Contributors

Experimental work was done by M.D. Martin. T.A. Hopp assisted M.D. Martin with characterization of the MTA1 antibody. Statistical analyses were done by G.M. Clark and S.G. Hilsenbeck. S.K. Mohsin and D.C. Allred assessed tumor pathology and constructed tissue arrays, and supervised assessment of MTA1 immunohistochemistry. C.K. Osborne, G.M. Clark, and D.C. Allred assembled the breast tumor specimens and patient outcome data used in the study. M.D. Martin and P. O'Connell conducted the experimental studies described herein and authored successive drafts of the manuscript. D.C. Allred and P. O'Connell designed and supervised the overall study.

Deleted: assisted with patholog

Deleted: ical

Deleted: of

Deleted: set up

Deleted: bank

Deleted: of patient

Deleted: samples

Deleted: designed

Deleted: s,

Deleted: wrote the

Deleted:

Deleted: initiated and

Deleted: sed

Deleted: entire

Deleted: have

#### Conflict of Interest Statement

M.D. Martin, G.M. Clark, D.C. Allred, and P. O'Connell, filed a US patent for measurement of MTA1 expression in breast cancer.

#### Acknowledgements

The authors thank Drs Gary Chamness and Jenny Chang for critical reading of the manuscript. This report was supported by grants CA30195 and CA58183 from the National Cancer Institute, National Institutes of Health, Department of Human Services, USAMRC Breast Cancer Training Grant DAMD-17-99-1-9047873, and USAMRC Breast Cancer "Concept" awards DAMD17-01-1-0478 and DAMD17-03-1-0648. This study was conducted under the aegis of Baylor College of Medicine Institutional Review Board Protocol 47830.

Deleted: done under the approved

Deleted: RB

Deleted: p

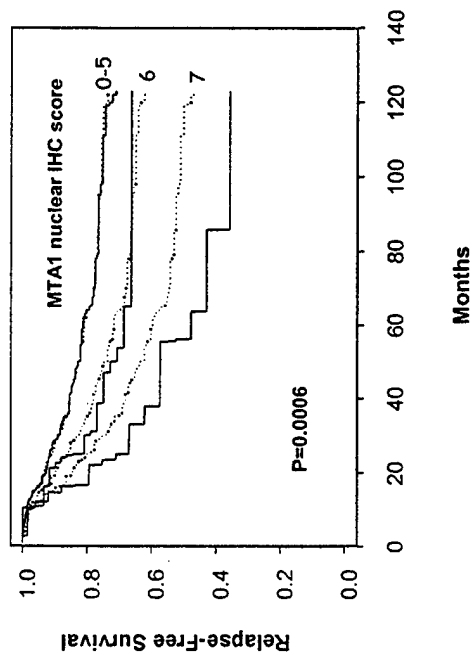
Deleted: BCM IRB

#### Reference List

1. Martin MD, Fischbach K, Osborne CK, Mohsin SK, Allred DC, O'Connell P. Loss of heterozygosity events impeding breast cancer metastasis contain the MTA1 gene. Cancer Res. 61:3578-80, 2001

2. Kumar R, Wang RA, Mazumdar A, Talukder AH, Mandal M, Yang Z et al. A naturally occurring MTA1 variant sequesters oestrogen receptor-alpha in the cytoplasm. *Nature* 2002;418 (6898) :654-7.
3. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod.Pathol.* 1998;11 (2) :155-68.
4. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* 2003;362 (9381) :362-9.
5. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model.* New York, NY: Springer-Verlag. 2000.





**Figure 1. Kaplan-Meier Curves.** Estimates of relapse-free survival and corresponding Cox regression estimates are shown for various values of MTA1 in node-negative untreated subjects (N=397). Solid lines represent Kaplan-Meier curves, and dashed lines represent Cox regression estimates. MTA1 values of 0-5, 6, 7, and 8 were coded as 0, 1, 2, and 3 respectively in the analysis. Kaplan-Meier estimates and log rank tests were used to display and test the univariate association between RFS or OS and MTA1. Cox proportional hazards regression was used to test the independent contribution of MTA1 after accounting for other potentially important covariates. Adjusted survival curves were generated using Cox regression estimates for various values of MTA1, with cohort averages being used for other covariates in the model. Plots have been truncated at 120 months for graphical presentation, but all data were included in the analyses. Analyses were performed using the SAS (Version 8.2, Cary NC), and Splus (Version 6.1, Insightful, Seattle, WA).

Deleted: Product Limits

Deleted: analysis

Deleted: Kaplan

Table 1. Analysis<sup>1</sup> of MTA1 Overexpression, Relapse-Free, and Overall Survival

Variable <sup>4</sup>	Relapse-Free Survival <sup>2</sup>		HazardRatio (95% CI)	Overall Survival <sup>3</sup>		HazardRatio (95% CI)
	Univariate P value	Multivariate P val		Univariate P Value	Multivariate P Value	
All Patients (N=1017; 326 recurrences; 414 deaths)						
MTA1_cut6	0.17	0.0015	1.91 (1.26-2.88)	0.76	0.415	NA
Positive LN	<0.0001	<0.0001	2.10 (1.80-2.46)	<0.0001	<0.0001	1.85 (1.62-2.01)
S-phase Fxn	0.0001	0.0435	1.17 (1.00-1.35)	<0.0001	0.189	NA
Tumour Size	<0.0001	0.0002	1.33 (1.22-1.59)	<0.0001	<0.0001	1.38 (1.18-1.60)
ER+	<0.0001	0.0009	0.65 (0.51-0.84)	0.0008	0.0240	0.74 (0.58-0.96)
PGR	0.0006	0.1917		<0.0001	0.012	0.74 (0.58-0.94)
Untreated Node-Negative Patients (N=394; 102 recurrences; 138 deaths)						
MTA1_cut6	0.0001	0.0006	2.68 (1.53-4.69)	0.139	0.11	
S-phase Fraction	0.007	0.072	1.26(0.99-1.60)	0.36	0.47	
Tumour Size	0.024	0.039	1.41(1.02-1.95)	0.045	0.046	1.32(1.01-1.74)
ERa	0.55	0.541		0.367	0.61	
PGR	0.93	0.303		0.455	0.57	
Treated Node Negative Patients (N=217; 33 recurrences; 34 deaths)						
MTA1_cut6	0.75	0.61		0.36		
S-phase Fraction	0.035	0.039	1.58(1.02-2.42)	0.20		
Tumour Size	0.86	0.78		0.57		
ERa	0.098	0.37		0.54		
PGR	0.099	0.26		0.11		

<sup>1</sup>Statistical methodology is reviewed in (5), hazard ratios are shown for a one level increase in the associated covariate, relative to baseline

<sup>2</sup>Relapse-free survival (RFS) was defined as the time from diagnosis to first recurrence (local or distant), or to last contact or death (censored).

<sup>3</sup>Overall survival (OS) was defined as the time from diagnosis to death from any cause or last contact (censored).

<sup>4</sup>Analysis of functional form (5) suggested that MTA1 scores above 5 are associated with proportionately increasing risk, while scores of 5 or less carry essentially the same risk. Immunostained slides were evaluated for the presence nuclear staining, and scored as previously described (3). S-phase (<6%, 6-10%, and >10%), and tumour size (<2 cm, 2-5 cm, >5 cm) and were trichotomized, and coded 0, 1, and 2, respectively. ER (ER-6F11, Novocastra, Newcastle upon Tyne, UK) and PgR (PgR.1294, Dako, Carpinteria, CA) were previously analyzed by immunohistochemistry using standard protocols

## Mechanisms of disease

# Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer

Jenny C Chang, Eric C Wooten, Anna Tsimelzon, Susan G Hilsenbeck, M Carolina Gutierrez, Richard Elledge, Syed Mohsin, C Kent Osborne, Gary C Chamness, D Craig Allred, and Peter O'Connell

## Summary

**Background** Systemic chemotherapy for operable breast cancer substantially decreases the risk of death. Patients often have de novo resistance or incomplete response to docetaxel, one of the most active agents in this disease. We postulated that gene expression profiles of the primary breast cancer can predict the response to docetaxel.

**Methods** We took core biopsy samples from primary breast tumours in 24 patients before treatment and then assessed tumour response to neoadjuvant docetaxel (four cycles, 100 mg/m<sup>2</sup> daily for 3 weeks) by cDNA analysis of RNA extracted from biopsy samples using HgU95 GeneChip.

**Findings** From the core biopsy samples, we extracted sufficient total RNA (3–6 µg) for cDNA array analysis using HgU95-Av2 GeneChip. Differential patterns of expression of 92 genes correlated with docetaxel response ( $p=0.001$ ). Sensitive tumours had higher expression of genes involved in cell cycle, cytoskeleton, adhesion, protein transport, protein modification, transcription, and stress or apoptosis; whereas resistant tumours showed increased expression of some transcriptional and signal transduction genes. In leave-one-out cross-validation analysis, ten of 11 sensitive tumours (90% specificity) and 11 of 13 resistant tumours (85% sensitivity) were correctly classified, with an accuracy of 88%. This 92-gene predictor had positive and negative predictive values of 92% and 83%, respectively. Correlation between RNA expression measured by the arrays and semiquantitative RT-PCR was also ascertained, and our results were validated in an independent set of six patients.

**Interpretation** If validated, these molecular profiles could allow development of a clinical test for docetaxel sensitivity, thus reducing unnecessary treatment for women with breast cancer.

*Lancet* 2003; **362**: 280–87

See Commentary

Breast Center and the Departments of Medicine, Pathology, and Molecular and Cellular Biology, Baylor College of Medicine, and the Methodist Hospital, Houston, TX USA (J C Chang MD, EC Wooten PhD, A Tsimelzon PhD, Prof S G Hilsenbeck PhD, M C Gutierrez MD, R Elledge MD, S Mohsin MD, Prof C K Osborne MD, Prof G C Chamness PhD, Prof D C Allred MD, Prof P O'Connell PhD)

**Correspondence to:** Dr Jenny Chang, Breast Center, Baylor College of Medicine, One Baylor Plaza, BCM 600, Houston, TX 77030, USA (e-mail: jchang@breastcenter.tmc.edu)

## Introduction

Adjuvant systemic treatment after surgery for breast cancer is the most crucial factor in reducing mortality—both chemotherapy and hormonal treatment reduce the risk of death in such patients.<sup>1–4</sup> However, although oestrogen-receptor status is predictive of response to hormonal treatments, there are no clinically useful predictive markers of a patient's response to chemotherapy. Therefore, all patients who are eligible for chemotherapy receive the same treatment, even though de novo drug resistance will result in treatment failures in many. The taxanes, docetaxel and paclitaxel, are a new class of antimicrotubule agent that are more effective than older drugs such as anthracyclines,<sup>5–7</sup> although results of clinical trials with taxanes and anthracyclines in combination show that only a small subset of patients benefit from the addition of taxanes.<sup>8,9</sup> There are no methods to distinguish between patients who are likely to respond to taxanes and those who are not. In view of the accepted practice of giving adjuvant treatment to most patients, even if the average expected benefit is low, the a priori selection of appropriate patients most likely to benefit from adjuvant treatment with taxanes would be a great advance in the clinical management of breast cancer.<sup>10</sup> A major impediment to study of predictors of effectiveness of adjuvant treatment is the absence of surrogate markers for survival and, consequently, large numbers of patients and long-term follow-up are needed.

We aimed to identify gene expression patterns in primary breast-cancer specimens that might predict response to taxanes. Neoadjuvant chemotherapy (ie, treatment before primary surgery) allows for sampling of the primary tumour for gene expression analysis, and for direct assessment of response to chemotherapy by monitoring changes in tumour size during the first few months of treatment.<sup>10,11</sup> Clinical response of the tumour to neoadjuvant chemotherapy is a valid surrogate marker of survival: patients whose tumours regress substantially after neoadjuvant chemotherapy have better outcome than those with modest response or clinically obvious disease that is resistant to chemotherapy.<sup>10,11</sup> With the advent of high-throughput quantification of gene expression, simultaneous assessment of thousands of genes is now possible, which allows identification of expression patterns in different breast cancers that might correlate with, and thereby predict, excellent clinical response to treatment.<sup>12–16</sup> These profiles have potential to explain the genetic heterogeneity of breast cancer and allow treatment strategies to be planned in accordance with their probability of success in individual patients. Hence, neoadjuvant chemotherapy provides an ideal platform from which to discover predictive markers of chemotherapy response. In our study, we took core needle biopsy samples of the primary breast cancer for gene expression profiling before patients received neoadjuvant docetaxel. We aimed first, to show that sufficient RNA

**GLOSSARY****ANEUPLOIDY**

Cells containing an abnormal complement of chromosomes.

**APOPTOSIS**

Programmed cell death. A genetic mechanism leading to induced cell death that involves activation of a cascade of genes. Apoptosis arises in normal tissue and can be associated with particular disease states.

**RESUBSTITUTION ESTIMATES**

Application of the classifier to the samples used to create it.

could be obtained from core biopsy samples to assess gene expression; second, to identify groups of genes that could be used to distinguish primary breast cancers that are responsive or resistant to docetaxel chemotherapy; and third, to identify gene pathways that could be important in the mechanism of resistance to docetaxel.

**Methods****Patients**

From September, 1999, to June, 2001, patients with locally advanced breast cancer (ie, primary cancers >4 cm, or clinically evident axillary metastases) were considered for a phase 2 study with neoadjuvant docetaxel. Inclusion criteria were (1) age greater than 18 years and a diagnosis of breast cancer confirmed by analysis of a core needle biopsy sample, (2) premenopausal status accompanied by appropriate contraception, (3) adequate performance status, and (4) adequate liver and kidney function tests (all within 1.5 times the institution's upper limit of normal). Patients were excluded if they had severe underlying chronic illness or disease, or were taking other chemotherapeutic drugs while on study.

This study (protocol H8448) was approved by the institutional review board of Baylor College of Medicine, Houston, TX, USA. Patients gave written informed consent.

**Clinical procedures**

We recorded clinical staging and size of primary tumour at the start of treatment, at every cycle, and after completion of four cycles of chemotherapy. Tumour size (product of the two largest perpendicular diameters) measured before and after four cycles of neoadjuvant chemotherapy was used to calculate the percentage of residual disease. The median residual disease was then calculated, and this degree of response was used to divide the cancers into two roughly equal groups—sensitive and resistant tumours—before we did gene expression analysis.

Before docetaxel was given, we did core biopsies of the primary cancers. To obtain sufficient tissue, we did about six core biopsies from every patient using an MC1410 MaxCore biopsy instrument (Bard, Covington, GA, USA). Samples were taken after patients had been given local anesthesia with the same entry point, but reorienting the needle. Two to three core biopsy specimens were immediately transferred for snap freezing at  $-80^{\circ}\text{C}$  for cDNA array analysis. The remaining specimens were fixed in formalin for diagnostic analysis and possible immunohistochemical analysis.

Four cycles of docetaxel were given at  $100\text{ mg/m}^2$  every 3 weeks, and we assessed clinical response after the fourth cycle, at 12 weeks. As part of standard care, patients were continued on neoadjuvant chemotherapy through the full four cycles, unless there was clear documentation of progressive disease, which we defined as increase in tumour size of more than 25%. After the course of

neoadjuvant docetaxel was complete, primary surgery was done and standard adjuvant treatment was given.

**RNA extraction and amplification**

We isolated total RNA from the frozen core biopsy specimens in accordance with protocols recommended by Affymetrix (Santa Clara, CA, USA) for GeneChip experiments. Total RNA was isolated with Trizol reagent (Invitrogen Corporation, Carlsbad, CA). Samples were subsequently passed over a Qiagen RNeasy column (Qiagen, Valencia, CA) for control of small fragments that affect RT-reaction and hybridisation quality (ECW, unpublished data). Each core biopsy yielded 3–6  $\mu\text{g}$  of total RNA. After RNA recovery, double-stranded cDNA was then synthesised by a chimeric oligonucleotide with an oligo-dT and a T7 RNA polymerase promoter at a concentration of 100 pmol/ $\mu\text{L}$ .

We did reverse transcription in accordance with protocols recommended by Affymetrix using commercially available buffers and proteins (Invitrogen Corporation). Biotin labelling and about 250-fold linear amplification followed phenol-chloroform clean up of the reverse-transcription reaction product and was done by in-vitro transcription (Enzo Biochem, New York, NY, USA) over a reaction time of 8 h. From each biopsy specimen, we hybridised 15  $\mu\text{g}$  of labelled cRNA onto the U95Av2 GeneChip using recommended procedures for prehybridisation, hybridisation, washing, and staining with streptavidin-phycoerythrin (SA-PE). Antibody amplification was done with a biotin-linked antibody to streptavidin (Vector Laboratories, Burlingame, CA) with a goat-IgG blocking antibody (Sigma, St Louis, MO, USA). A second application of the SA-PE dye was used after additional wash steps had been done. After automated staining and wash protocols (Affymetrix protocol EukGE-2v4), the arrays were scanned by the Affymetrix GeneChip scanner (Agilent, Palo Alto, CA) and quantitated with Micoarray suite version 5.0 (Affymetrix). The U95Av2 GeneChip consists of about 12 625 probe sets, each containing about 16 perfect match and corresponding mismatch 25mer oligonucleotide probes representing sequences (genes), most of which have been characterised in terms of function or disease association. The raw, un-normalised probe level data were then analysed by dChip (<http://dchip.org>) for final normalisation and modelling. Median intensity was used for the normalisation of the 24 arrays and the perfect match/mismatch (PM/MM) modelling algorithm was used.

**Semiquantitative RT-PCR**

We did semi-quantitative RT-PCR (sqRT-PCR) measurement of gene expression levels using the same amplified cRNA hybridised to the GeneChip. 20 genes were selected for analysis on the basis of their high variation in expression. Primers were designed for these loci with the sequences freely available from the Entrez Nucleotide database<sup>17</sup> and the Primer3 algorithm for primer design. Product sizes were kept short (<150 bp) to allow the maximum ability to work under varying conditions relative to cRNA quality. Primers were optimised with a reverse-transcribed mixture of six samples. 15 duplicate reactions were prepared and samples were taken at alternating cycle numbers between 15 and 33 to ensure that the sqRT-PCR reaction products were in a linear range of accumulation. These samples were then arranged in ascending order, diluted with 10  $\mu\text{L}$  loading buffer, and 3  $\mu\text{L}$  of each sample was loaded onto 6% denaturing acrylamide gels. Electrophoresis at 60 W was done for 2 h, or until sufficient

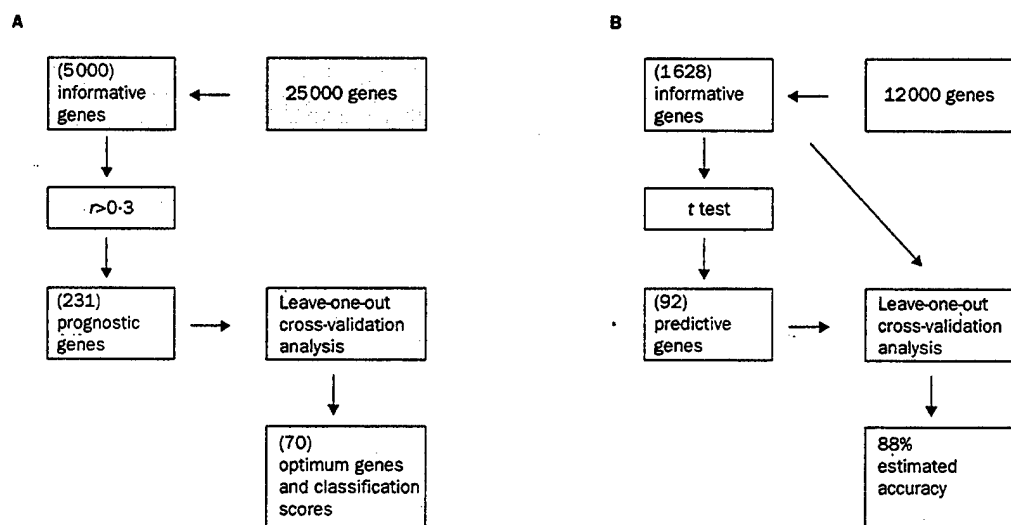


Figure 1: Two methods of statistical analysis

A: the prognostic analysis used by van't Veer and colleagues<sup>18</sup> used oligonucleotide microarrays with 25 000 genes, from which 5000 variably expressed genes were selected by filtering. Of these, 231 genes were significantly associated with prognostic outcome ( $r > 0.3$ ). These 231 genes were then rank-ordered on the basis of the magnitude of the correlation coefficient and selected in groups of five to construct the smallest optimum classifier. Leave-one-out analysis was then done with 231 genes that were correlated with outcome to select a classification set of 70 genes. B: statistical analysis methods used in this study: a subset of 1628 genes was selected by filtering on signal intensity to eliminate genes with uniformly low expression or genes whose expression did not vary significantly across the samples.

separation of the xylene cyanol and bromophenol blue dyes was achieved. Gels were then fixed, removed from the rear plate, transferred to filter paper, and dried. We first assessed these dry gels using autoradiography (about 8 h exposure, no intensification), and analysable gels were then exposed to phosphorimaging screens. Primers that failed to produce a single clear band were attempted again with different annealing temperatures until a single band was produced.

15 of the 20 primers chosen proved suitable to use and gave clean, single bands for analysis. The remaining five failed to optimise properly and were not included in any further analysis. Although high-cycle samples inevitably achieved pixel-saturation, care was taken to keep exposure times to a minimum, so as to keep intensity within the informative range on most cycle-totals within each set. To determine the linear range of the 15 primers, we analysed their absolute intensities using Microsoft Excel graphing functions. We then did phosphorimager quantification analysis (Bio-Rad Laboratories, Hercules, CA), and RT-PCR product band intensities were quantitatively compared with normalised, model-based estimates of expression from the GeneChip data.

#### Statistical analysis

The analytical approach used in this study (figure 1) was similar to the successful methods described previously.<sup>18</sup> After scanning and low-level quantification using Microarray Suite (Affymetrix), we used DNA-Chip analyser dchip version 1.2 to adjust arrays to a common baseline<sup>19</sup> and estimated expression using Li and colleagues' PM-MM model.<sup>20,21</sup> We eliminated genes that were not present in at least 30% of samples, and exported expression data for the remaining 6849 genes to BRB Arraytools version 2.1c<sup>22</sup> for more filtering and analysis. After transforming all data by taking logarithms, we ranked genes by variability over all 24 samples, and we retained the 1628 genes that were significantly more variable than the median variance.

We selected differentially expressed genes from the filtered gene list using the two-sample  $t$  test, and then used a global permutation test as an overall, multiple comparison-free test of whether the number of differentially expressed genes exceeded that which might arise by chance. In this test, the observed number of significantly differentially expressed genes was compared with the distribution of numbers of differentially expressed genes generated by repeatedly PERMUTATING the labels of the samples and recalculating the  $t$  test at the specified level of significance.

Next, we developed a classifier to predict response. With a list of discriminatory genes and their associated  $t$  values, we used the compound covariate predictor method of Radmacher and colleagues<sup>23</sup> to construct a linear classifier. RESUBSTITUTION ESTIMATES of classification success, in which the classifier is applied to the same samples used to create it, are invariably biased (ie, they are overly optimistic).<sup>24,25</sup> Therefore, we used an external cross-validation procedure to generate a less biased estimate of classification success. Starting with 1628 genes that had significant variation in expression, and which were filtered without any respect to class membership, the entire gene selection and classifier construction process was repeated in a leave-one-out cross-validation to estimate classifier performance. Finally, to assess whether the degree of successful classification we noted could have arisen by chance, the entire cross-validation procedure was repeated 2000 times, permutating the sample labels every time. The observed cross-validated classification success rate was then compared with the distribution of classification success in the permutation analysis. Cross-validated performance was summarised by observed sensitivity and specificity, and associated exact binomial confidence intervals. Resubstitution classifier values were also used to generate a receiver operating characteristic curve (ROC curve) and to estimate the area under the curve.

Patient	Age (years)	Menopausal status	Ethnic origin	Unidimensional tumour size (cm)	Clinical axillary nodes	Oestrogen-receptor status	Progesterone-receptor status	HER-2	Tumour type
1	37	Premenopausal	Hispanic	10×10	No	-	-	-	IMC
2	55	Postmenopausal	Hispanic	10×8	Yes	-	-	+	IDC
3	41	Premenopausal	Black	6×5	Yes	+	+	-	IDC
4	43	Premenopausal	Black	15×13	Yes	+	-	-	IMC
5	50	Postmenopausal	Black	20×23	Yes	-	-	-	IDC
6	55	Postmenopausal	Black	11×11	Yes	+	+	-	IDC
7	42	Premenopausal	Black	7×9	Yes	+	+	-	IMC
8	63	Premenopausal	Black	7×8	Yes	+	+	-	IMC
9	50	Postmenopausal	Black	13×9	No	+	+	-	IDC
10	38	Premenopausal	Hispanic	8×8	Yes	+	+	-	IMC
11	58	Postmenopausal	Hispanic	7×7	Yes	+	+	-	IMC
12	62	Postmenopausal	Hispanic	4×4	Yes	+	-	-	IDC
13	40	Premenopausal	Hispanic	5.5×4.5	No	+	+	-	IMC
14	36	Premenopausal	Black	6×6	Yes	+	+	-	IDC
15	56	Postmenopausal	Black	5×5.5	No	+	-	-	IMC
16	38	Premenopausal	White	6×6	Yes	+	-	-	IDC
17	54	Postmenopausal	White	5×6	Yes	+	+	+	IDC
18	52	Postmenopausal	White	10×10	No	+	+	-	IDC
19	57	Postmenopausal	White	8×8	No	-	-	-	IDC
20	52	Postmenopausal	Black	10×10	No	-	-	-	IDC
21	44	Premenopausal	Black	11×11	No	-	-	-	IDC
22	41	Premenopausal	Black	6×5	Yes	+	+	-	IDC
23	38	Premenopausal	White	8×8	Yes	+	+	-	IDC
24	54	Postmenopausal	Black	9×7	No	+	+	-	IDC

HER-2=HER-2/neu oncogene detected by immunohistochemical analysis. -negative. +=positive. IMC=invasive mammary carcinoma. IDC=invasive ductal carcinoma.

Table 1: Characteristics of patients in the training set

The classifier was partly validated with an independent set of six patients treated in the same clinical trial as those in the training set. RNA was obtained from pretreatment biopsy samples and hybridised to HgU95av2 GeneChips exactly as described for the training sample. Probe level data were adjusted to the same baseline array as the training set, and gene expression values were calculated with previously estimated probe sensitivity values derived from the training sample. The 92-gene classifier was then applied to predict response in every new sample.

#### Role of the funding source

The study sponsors did not contribute to the study design, collection, analysis, or interpretation of data. The manuscript was reviewed with only minor editorial changes by one of the study's sponsors, Aventis Pharmaceutical.

## Results

### Assessment of clinical response

We included 24 patients, and their clinical characteristics are shown in table 1. Unidimensional median tumour size before treatment was 8 cm (range 4–30 cm). Before doing gene expression analysis, we defined tumour sensitivity and resistance on the basis of the percentage of residual disease after treatment. We first determined that the median residual disease after chemotherapy was 30%. We then arbitrarily defined sensitive tumours as those that had 25% or less residual disease, and resistant tumours as those with more than 25% residual disease, since this cutoff divides the patients into two almost equally sized groups for statistical comparison. In this study of locally advanced breast cancer, tumours were large and a regression of at least 75% after chemotherapy would almost certainly represent a clinically important response. Of these 24 patients, 11 (46%) were sensitive to docetaxel and 13 (54%) were resistant. Of the sensitive tumours, five patients (45%) had minimal residual disease (<10% residual tumour), whereas of the resistant tumours, seven (58%) had residual tumour mass of 60% or greater, and three (23%) of these residual tumours were 100% or greater of baseline.

### Selection of discriminatory genes

To select discriminatory genes, we compared expression data in the sensitive and the resistant tumours (figure 2). First, we selected a subset of candidate genes by filtering on signal intensity to eliminate genes with uniformly low expression or genes whose expression did not vary significantly across the samples, retaining 1628 genes. After log transformation, a *t* test was used to select discriminatory genes. *t* tests with nominal *p* values of 0.001, 0.01, and 0.05 selected 92, 300, and 551 genes, respectively, for which expression differed in sensitive and resistant groups—ie, differentially expressed. The probability that these numbers of genes would be selected by chance alone was estimated to be 0.0015, 0.001, and less than 0.001 respectively (table 2). These results can be reviewed with data at the gene expression omnibus.<sup>26</sup>

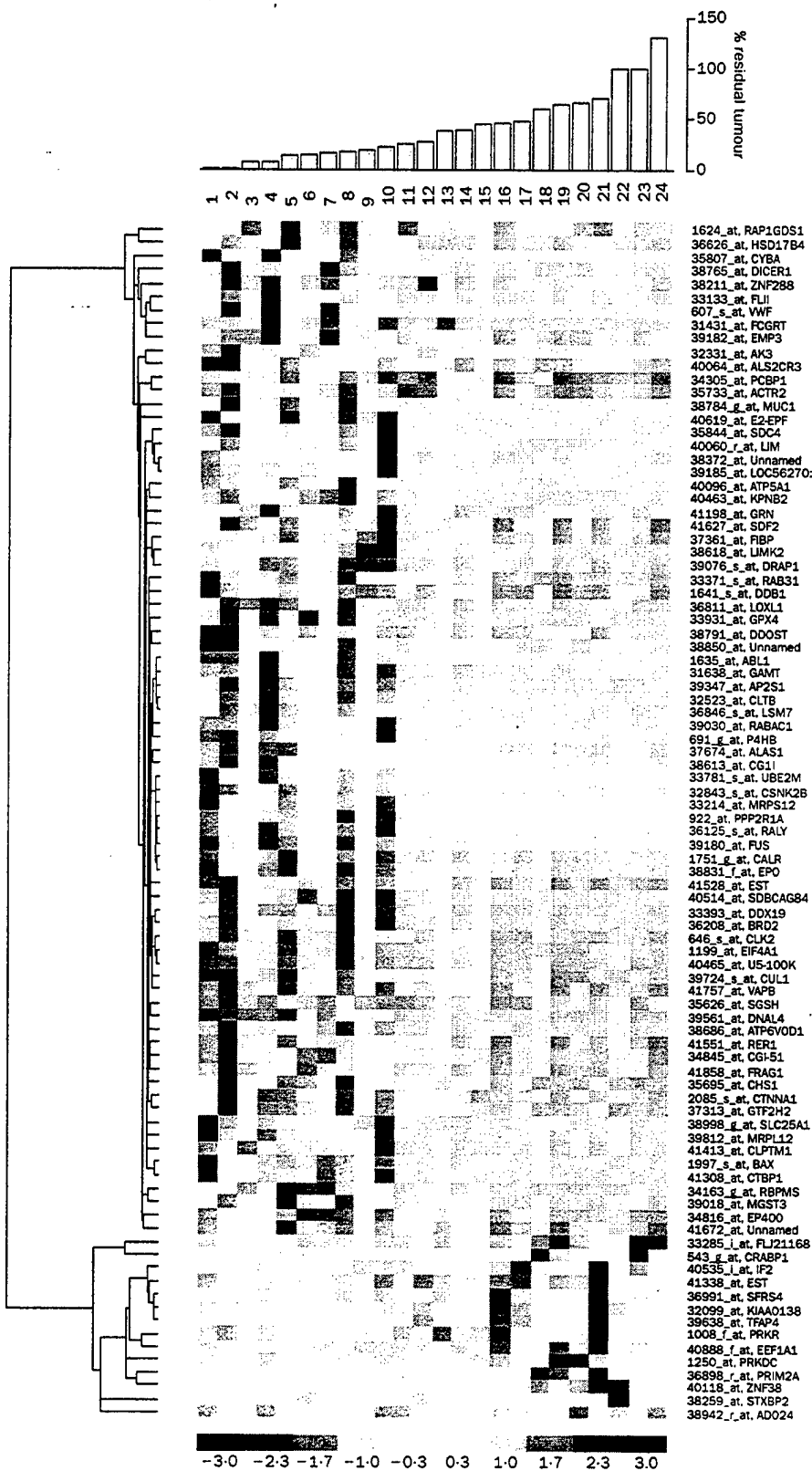
### Functional classification of discriminatory genes

The 92 genes classed as most significantly "differentially expressed" at *p*=0.001 are listed in the webtable (<http://image.thelancet.com/extras/01art11086webtable.pdf>) (figure 2). These genes showed 4.2–2.6-fold decreases or 2.5–15.7-fold increases in expression in resistant compared with sensitive tumours. Functional classes of these differentially expressed genes included stress or apoptosis (21%), cell adhesion or cytoskeleton (16%), protein transport (13%), signal transduction (12%), RNA transcription (10%), RNA splicing or transport (9%), cell cycle (7%), and protein translation (3%); the remainder (9%) have unknown functions. 14 of these 92 genes were overexpressed in the treatment-resistant cluster with major categories including unknown function, protein

	p value for gene selection		
	0.001	0.01	0.05
Number of differentially expressed genes	92	300	551
Permutation <i>p</i> *	0.0015	0.001	0.001

\*The proportion of permutations in which the number of genes selected exceeds the observed number of genes.

Table 2: Group comparison analysis, with different nominal *p*-values (0.001, 0.01, 0.05)



**Figure 2: Hierarchical clustering of genes correlated with docetaxel response**  
 Expression levels above the mean for the gene are shown in red and expression levels below the mean for the gene are shown in blue. The colour scale ranges from 3 SDs or more below the mean (darkest blue) to 3 SDs above the mean (darkest red). Affymetrix probe set identifiers and gene symbols are shown on the right-hand side of the figure.

translation, cell cycle, and RNA transcription. Tubulin isoforms were associated with docetaxel resistance.

Of the 78 genes overexpressed in docetaxel-sensitive tumours, major categories were stress or apoptosis, adhesion or cytoskeleton (no genes with this function were overexpressed in resistant tumours), protein transport, signal transduction, and RNA splicing or transport. In sensitive tumours, genes involved in apoptosis (eg, overexpression of *BAX*, *UBE2M*, *UBCH10*, *CUL1*), and DNA damage-related gene expression (eg, overexpression of *CSNK2B*, *DDB1*, and *ABL1*, and underexpression of *PRKDC*) seem to contribute to docetaxel sensitivity.

#### Leave-one-out cross-validation

In this cross-validation analysis, we began with all 1628 filtered genes to avoid selection bias.<sup>24,25</sup> Every observation in turn was left out and the remaining samples were used to select differentially expressed genes; we then constructed a compound covariate predictor to classify the left-out sample. Ten of 11 sensitive tumours (91% specificity, [95%CI 0.59–1.00]) and 11 of 13 resistant tumours (85% sensitivity [0.55–0.98]) were correctly classified, for an overall accuracy of 88% (68–97%). Results of permutation testing showed that such a high cross-validated classification accuracy is significant ( $p=0.008$ ). The analogous predictor, constructed with 92 genes selected with use of all 24 samples, yielded identical classification success. With this predictor, positive and negative predictive values for response to docetaxel were 92% and 83%, respectively, and the area under the ordinary receiver operating characteristic (ROC) curve was 0.96 (figure 3).

#### Confirmation of expression measurements

To confirm measurement of RNA concentrations, expression values derived from adjusted Affymetrix data were correlated with values from sqRT-PCR for 15 variably expressed genes (table 3). Spearman rank correlations were positive for 13 genes and significantly positive for six of 15 genes.

#### Validation in an independent cohort

The six additional patients enrolled in this prospective clinical study were studied to partly validate the 92-gene

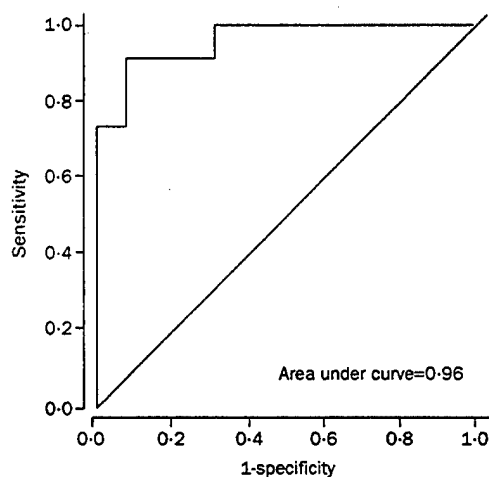


Figure 3: Receiver operating characteristic (ROC) curve for predicting response to docetaxel

Affymetrix probe set	Number	Pearson correlation		Spearman rank correlation	
		r	p	r <sub>s</sub>	p
ACTB 32318_s_at	5	0.81	0.09	0.90	0.04
ATP6VOE 33875_at	5	0.28	0.65	0.10	0.87
BMF1 1728_at	8	0.90	0.002	0.21	0.61
CALM3 1158_s_at	7	0.52	0.23	0.64	0.12
FUCA1 41814_at	6	0.77	0.07	0.94	0.00
GLRX 34311_at	8	0.74	0.03	0.50	0.21
IFITM1 676_g_at	5	0.74	0.15	0.70	0.19
LAMR1 256_s_at	8	0.69	0.06	0.85	0.01
LMNA 37378_r_at	5	-0.08	0.90	-0.40	0.50
MUC1 38783_at	8	0.84	0.01	0.71	0.05
MYO10 35362_at	8	0.15	0.72	0.05	0.91
PLD 36184_at	4	-0.41	0.59	-0.80	0.20
PSMD5 32240_at	8	0.27	0.52	0.33	0.42
SERPINB5 863_g_at	8	0.75	0.03	0.81	0.01
SPARCL1 36627_at	6	0.92	0.01	1.00	0.00

Table 3: Correlation of Affymetrix expression data with sqRT-PCR derived values.

Correlations positive for 13 genes and significantly positive for 6 of 15 genes

predictive classifier. In this small set, all six patients had sensitive tumours and were correctly classified by our predictive method.

#### Discussion

We obtained sufficient RNA from small core biopsy samples of human breast cancers, to assess patterns of gene expression in individual tumours and identified molecular profiles using gene expression patterns of human primary breast cancers to accurately predict sensitivity to docetaxel in women with primary breast cancer.

Gene expression patterns associated with docetaxel sensitivity and resistance are highly complex. In the past, investigators using single gene biomarkers to assess sensitivity and resistance to chemotherapy have seldom produced conclusive results. For example, in a breast cancer study the researchers did not note any correlation between commonly measured predictive and prognostic markers (HER-2, p53, p27, or epidermal growth factor receptor) and taxane sensitivity.<sup>27</sup> Reports of different cancer types have suggested that alterations in expression levels of  $\beta$  tubulin isoforms might represent an important and complex mechanism of taxane resistance.<sup>28</sup> We noted that overexpression of some  $\beta$  tubulin isoforms was associated with docetaxel resistance in some tumours, but not all. These results suggest that the patterns of gene expression for sensitivity and resistance are likely to involve multiple gene pathways, and that integration of many genes in these pathways leads to drug sensitivity and resistance. Our results lend support to the idea that assessment of expression of a few individual genes will not be powerful enough to untangle the heterogeneity of clinical breast cancers, but that patterns of expression of many genes could be successful in distinguishing between sensitive and resistant tumours.

A key point of this study was to focus on genes that could be reliably measured and to exclude those that were unlikely to be expressed in any sample. We did not design this study to discover specific genes for docetaxel response or resistance, but rather to identify patterns of many genes that could be used as a predictive test in patients with breast cancer. As a result, our analysis will have excluded some differential genes with low expression, some of which might be biologically interesting. For example, that spindle checkpoint dysfunction is an important cause of aneuploidy in human cancers has been suggested. The



serine-threonine kinase gene *STK6* (AURORA A)<sup>28</sup> might constitute a mechanism of spindle checkpoint dysregulation, and its amplification has been shown to predict resistance to taxanes.<sup>29</sup> Indeed, we did note differential expression between sensitive and resistant tumours—overexpression of *STK6* was about 1.4-fold higher in docetaxel-resistant tumours than in those that were sensitive to the drug (mean expression 506 and 695 in sensitive and resistant tumours, respectively;  $p=0.046$ ). Nevertheless, this gene was not part of the 92-gene classifying list because of its overall low expression. This classifying list does not include all genes relevant to docetaxel sensitivity and resistance, but rather, identifies patterns of many genes that could be used as a predictive clinical test.

There is little information about the usefulness of gene expression arrays in human breast cancers.<sup>18,30-32</sup> Van't Veer and colleagues,<sup>18</sup> using printed oligonucleotide microarrays, noted that gene expression profiles were more accurate predictors of outcome in a small set of 78 young women with node-negative breast cancer than standard clinical and histological criteria. The same investigators subsequently validated this 70-gene classifier in a cohort of 295 patients, many of whom were not in the original study.<sup>31</sup> The signature of poor prognosis included genes regulating cell cycle, invasion, metastasis, and angiogenesis. Perou and colleagues<sup>32</sup> and Sorlie and colleagues<sup>33</sup> used cDNA arrays and identified distinct patterns of gene expression that were termed basal or luminal. These groups differed from each other with respect to clinical outcome.<sup>18,31</sup> Unlike these earlier publications that dealt with patient prognosis, our aim was to identify gene expression patterns that could predict response or resistance to docetaxel in patients with primary breast cancer.

Although breast cancers are highly heterogeneous, the classifying gene list gives some clues to the mechanisms of sensitivity and resistance in some tumours. In general, resistant tumours overexpressed genes associated with protein translation, cell cycle, and RNA transcription functions, whereas sensitive tumours overexpressed genes involved in stressor apoptosis, cytoskeleton, adhesion, protein transport, signal transduction, and RNA splicing or transport. Consistent with an apoptosis-induction mode of action for taxanes, sensitive tumours had higher expression of apoptosis-related proteins (eg, BAX, UBE2M, UBCH10, CUL1). DNA damage-related gene expression in docetaxel-sensitive tumours (overexpression of *CSNK2B*, *DDB1*, *ABL*, and underexpression of *PRKDC*) also seems to contribute to docetaxel sensitivity.

Furthermore, in sensitive tumours, overexpression of genes implicated in stress-related pathways was also noted, especially heat shock proteins. Overexpression of heat shock protein 27 (HSP27) has been associated with doxorubicin resistance in the MDA-MB-231 breast cancer cell line.<sup>34</sup> By contrast, the same investigators have shown that HSP27-overexpressing cell lines remain sensitive to docetaxel (Fuqua S, personal communication), suggesting that different non-cross-resistant agents could have different gene patterns of sensitivity and resistance. If true, then specific patterns of gene expression could be used as tools to choose between doxorubicin and docetaxel.

In a leave-one-out cross-validation procedure, the classifier that included genes selected at the nominal value of  $p \leq 0.001$  correctly classified tumours as sensitive or resistant in nearly 90% of cancers. Additionally, the predictive value of this classifier compares very favourably with that of oestrogen-receptor status, which is the only

validated factor that can predict response to hormone treatment in breast cancer. Oestrogen-receptor has a positive predictive value for response to hormone therapy of about 60%, and a negative predictive value of about 90%.<sup>34</sup> If about 70% of breast cancers are oestrogen-receptor positive, then sensitivity and specificity for hormone responsive and non-responsive tumours are about 93% and 50%, respectively, and the area under the ROC curve for oestrogen receptor is only about 0.72. The docetaxel classifier has positive and negative predictive values of 92% and 83%, respectively, and the area under the ROC curve of 0.96 (figure 3). Although these predictive values are likely to be slightly biased and have wide confidence intervals, these results suggest that classifiers based on gene expression would probably compare favourably with other clinically validated predictive markers.

Differences in RNA expression were confirmed by sqRT-PCR for a sample of genes. Furthermore, we have validated our classifier in an independent set of six consecutively treated patients, all of whom responded to treatment. Although the validation set is very small, it does lend support to the suggestion that gene expression arrays could be used to predict effectiveness of treatment.

This study shows that expression array technology can effectively and reproducibly classify tumours according to response or resistance to docetaxel chemotherapy. To ultimately define the molecular portrait of cancers sensitive or resistant to docetaxel, our results should be validated in a study with a large independent cohort of patients. Further patient recruitment and analysis will refine the gene list by which to classify tumours. This type of molecular profiling could have important clinical implications in defining the optimum treatment for an individual patient, thus reducing the use of unproductive treatments, unnecessary toxicity, and overall cost.

#### Contributors

J C Chang and E C Wooten contributed equally to this study. J C Chang designed the study, designed clinical experiments, gathered study samples, and wrote the manuscript. E C Wooten did laboratory experiments and contributed to the writing of the manuscript. A Tsimelzon and S G Hilsenbeck did statistical analysis, M C Gutierrez, S Mohsin, and D C Allred did pathological assessment of biopsy samples. C K Osborne and R Elledge did the clinical study and wrote the manuscript. G C Chamness and P O'Connell designed the experimental studies, and wrote the manuscript.

#### Conflict of interest statement

J C Chang, has received a Research Grant-in-Aid and is on the Speakers Bureau for Aventis. R Elledge has received a research grant from Aventis. J C Chang and P O'Connell have filed a US patent for docetaxel gene expression pattern.

#### Acknowledgments

We thank Mike T Lewis for his comments and manuscript review. This study was supported in part by the US Army Medical Research and Materiel Command DAMD17-01-0132, a Grant-in-Aid (US 11115) from Aventis Pharmaceutical, the Emma Jacobs Clinical Breast Cancer Fund, and the Breast Cancer SPORE, P50 CA50183 from the National Cancer Institute.

#### References

- 1 The Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women. *Lancet* 1992; 339: 1-15 (review).
- 2 The Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31 000 recurrences and 24 000 deaths among 75 000 women. *Lancet* 1992; 339: 71-85.
- 3 The Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 351: 1451-67.

- 4 The Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 352: 930-42.
- 5 Hortobagyi GN. Docetaxel in breast cancer and a rationale for combination therapy. *Oncology* 1997; 11: 11-15.
- 6 Chan S, Friedrichs K, Noel D, et al. Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer: the 303 Study Group. *J Clin Oncol* 1999; 17: 2341-54.
- 7 Nabholz JM, Patterson A, Dirix L, et al. A phase III trial comparing docetaxel (T), doxorubicin (A) and cyclophosphamide (C) (TAC) to (FAC) as first line chemotherapy for patients with metastatic breast cancer. *Proc Am Soc Clin Oncol* 2001; 20: 22a.
- 8 Henderson IC, Berry D A, Demetri GD, et al. Improved disease free survival and overall survival from the addition of sequential paclitaxel but not from escalation of doxorubicin in the adjuvant chemotherapy of patients with node-positive primary breast cancer. *Proc Am Soc Clin Oncol* 1998; 17: 101.
- 9 Aapro MS. Adjuvant therapy of primary breast cancer: a review of key findings from the 8th international conference, St Gallen. *Oncologist* 2001; 6: 376-85.
- 10 Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol* 1998; 16: 2672-85.
- 11 Mamounas EP. Preoperative doxorubicin plus cyclophosphamide followed by preoperative or postoperative docetaxel. *Oncology* 1997; 11(suppl 6): 37-40.
- 12 Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467-70.
- 13 Khan J, Simon R, Bittner M, et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Research* 1998; 58: 5009-13.
- 14 Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* 1996; 14: 1675-80.
- 15 Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR, Elkahoul AG. In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res* 1999; 59: 5656-61.
- 15 Perou CM, Jeffrey SS, van de Run M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA* 1999; 96: 9212-17.
- 16 Whitehead Institute/MIT centre for genome research. Primer 3. <http://www-genome.wi.mit.edu/cgi-bin/primer/primer3-www.cgi> (accessed May 17, 2003).
- 17 van't Veer LJ, Dai H, van De Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415: 530-36.
- 18 Schadt EE, Li C, Ellis B, Wong WH. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J Cell Biochem Suppl* 2001; 37 (suppl 37): 120-25.
- 20 Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* 2001; 98: 31-36.
- 21 Li C, Wong WH. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2001; 2: research0032.1-0032.11.
- 22 Simon R, Peng A. <http://linus.nci.nih.gov/BRB-ArrayTools.html> (accessed May 17, 2003).
- 23 Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol* 2002; 9: 505-11.
- 24 Ambrose C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci USA* 2002; 99: 6562-66.
- 25 Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* 2003; 95: 14-18.
- 26 NCBI. Gene expression omnibus: accession numbers GSE 349 (resistant samples) and GSE 350 (sensitive samples). [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accessed May 17, 2003).
- 27 Van Poznak C, Tan L, Panageas KS, et al. Assessment of molecular markers of clinical sensitivity to single-agent taxane therapy for metastatic breast cancer. *J Clin Oncol* 2002; 20: 2319-26.
- 28 Dumontet C, Sikic BI. Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *J Clin Oncol* 1999; 17: 1061-70.
- 29 Anand S, Penhryn-Lowe S, Venkitaraman AR. AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. *Cancer Cell* 2003; 3: 51-62.
- 30 van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347: 1999-2009.
- 31 Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; 98: 10869-74.
- 32 Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; 406: 747-52.
- 33 Hansen RK, Parra I, Lemieux P, Oesterreich S, Hilsenbeck SG, Fuqua SA. Hsp27 overexpression inhibits doxorubicin-induced apoptosis in human breast cancer cells. *Breast Cancer Res Treat* 1999; 56: 187-96.
- 34 Osborne CK, Yochmowitz MG, Knight WA, 3rd, McGuire WL. The value of estrogen and progesterone receptors in the treatment of breast cancer. *Cancer* 1980; 46 (suppl 12): 2884-88.